Supporting Information for

Side chain dynamics of carboxyl and carbonyl groups in the catalytic function of Escherichia coli ribonuclease H

Kate A. Stafford†, Fabien Ferrage‡, Jae-Hyun Cho§*, and Arthur G. Palmer, III†*

†Department of Biochemistry and Molecular Biophysics, Columbia University, 630 West 168th Street, New York, NY 10032 U.S.A.

‡Ecole Normale Supérieure, Département de Chimie, UMR 7203 CNRS-UPMC-ENS, 24, rue Lhomond, 75005 Paris, France.

§Department of Biochemistry and Biophysics, Texas A&M University, College Station, Texas 77843-2128, U.S.A.

Experimental Details

NMR experiments

(^{13}C, ^{15}N, 50% ^{2}H)-labeled Escherichia coli RNase H was dissolved in 99% $^{2}H_20$ (pHreading = 5.8, Mes 20mM, NaCl 50mM). All NMR experiments were conducted at 298K using a Bruker Avance 800 MHz spectrometer equipped with a cryogenic probe. Longitudinal (R_1), CSA/DD cross-correlation transverse (η_{xy}), rotating frame ($R_1\rho$) relaxation rate constants for 13Cγ/δ of side chain carbonyl/carboxyl groups were measured using the pulse sequence as described.[1] R_1 rate constants were measured with four relaxation delays; 128 (x2), 528 (x2), 968 (x2), and 1448 (x2) ms. η_{xy} were measured with 25 (x2) ms relaxation delay. For $R_1\rho$ rate constants, the spin-lock RF amplitude was $\omega_{SL}/2\pi$ = 1.1 kHz. The RF transmitter frequency for the spin-lock pulse was changed from 174 ppm to 184 ppm in 1 or 2 ppm steps to change effective magnetic field strength. The reported values are $R_1\rho$ values at 90° tilt angle. Relaxation delays were 10, 25, 40 and 80 ms. $R_1\rho$ rate constants for the Mg-bound RNase H were measured with 100mM $MgCl_2$.

Model free analysis of the 13Cγ/δ relaxation rates was performed using a simultaneous fit of order parameter (S^2) and the local correlation time (t_e) to the R_1 and η_{xy}. Errors of the fitting were estimated by Monte Carlo simulation using 300 random synthetic data sets. The isotropic rotational diffusion correlation time, τ_c, was used for model free analysis, although D_{par}/D_{per} = 1.24. It was shown that ignoring the modest anisotropy ($D_{par}/D_{per} < 1.3$) does not affect the model free analysis significantly.[2,3] The rotational correlation time of RNase H[4] was corrected

using Stokes' law to take into account the viscosity difference in 100% D_2O. Average CSA of backbone carbonyl 13C measured by solid state NMR[5] were used for Asn and Gln; σxx = 240.87 ppm, σyy = 196.62 ppm σzz = 93.5 ppm. For carboxyl 13C of Asp and Glu, CSA were assumed to be σxx = 242 ppm, σyy = 191 ppm σzz = 105 ppm.[6] The effect of choice of CSA values were tested with repeated model free analysis with extreme CSA values.

MD Simulation

Molecular dynamics simulations of E. coli RNase H were initiated from the structure 2RN2 and protonated in accordance with experimental pKa measurements to reflect pH values of 5.5 (D10 protonated) and 8.0 (D10 unprotonated). Using Maestro 8.5, crystallographic waters were removed, the structure was solvated in TIP3P water in a cubic box with a 10 Å buffer region from solute to box boundary, and the system was neutralized with Cl- ions. The protein was described with the Amber99SB force field.[7] Simulations were performed using Desmond Academic Release 3 or source release 2.4.2.1.[8] Electrostatics were calculated with the PME method. Simulations used a 2.5fs inner timestep on a 1-1-3 RESPA cycle and were performed in the NVT ensemble using a Nosé-Hoover thermostat after equilibration to constant box volume in the NPT ensemble. Simulation trajectories were calculated for a total of 100ns. Order parameters were calculated from trajectories using the equation[9]:

$$S^2 = \frac{1}{2}(3\sum_{i=1}^{3}\sum_{j=1}^{3} < \mu_i \mu_j >^2 - 1)$$

in which μ_i and μ_j represent the x, y, and z-components of a unit vector mu in the direction of a given chemical bond (C^{β}-C^{γ} for Asn and Asp, C^{γ}-C^{δ} for Gln and Glu). This calculation represents the long-time limit of the angular reorientational correlation function for a given bond vector. Order parameters were calculated as averages over discrete blocks of 10ns simulation time, a length chosen to correspond to the approximate time scale of global tumbling. Additional description of the low-pH trajectory is available elsewhere.[10]

Table S1. $^{13}C^{\gamma/\delta}$ R1ρ (θ = 90°) relaxation rate constants for apo and Mg^{2+} bound RNaseH

Residue	R1ρ	
	Apo	Mg^{2+}-Bound
10	23.91 ± 0.86	N.A.
16	15.66 ± 0.11	14.77 ± 0.11
32	19.41 ± 0.41	21.70 ± 0.27
44	12.67 ± 0.31	22.71 ± 0.17
48	22.42 ± 2.27	23.97 ± 9.93
57	25.15 ± 0.84	32.06 ± 2.19
70	16.44 ± 0.45	30.78 ± 0.33
72	5.86 ± 0.07	5.17 ± 0.10
76	11.11 ± 0.09	9.91 ± 0.26
84	12.61 ± 0.36	12.97 ± 0.01
94	11.08 ± 0.05	10.65 ± 0.01
100	47.38 ± 6.70	44.19 ± 6.20
102	28.83 ± 0.70	34.33 ± 0.39
105	17.37 ± 0.31	20.94 ± 1.09
113	5.15 ± 0.10	5.22 ± 0.16
115	4.23 ± 0.03	4.36 ± 0.01
119	10.42 ± 0.11	12.59 ± 0.01
129	21.22 ± 0.45	23.33 ± 0.59
131	8.68 ± 0.13	10.13 ± 0.01
134	17.31 ± 0.37	23.57 ± 0.30
135	6.81 ± 0.09	8.33 ± 0.01
143	6.95 ± 0.10	7.25 ± 0.03
147	3.55 ± 0.04	3.91 ± 0.02
148	31.97 ± 1.62	31.74 ± 0.53
152	7.54 ± 0.08	10.58 ± 0.05
154	3.02 ± 0.03	3.55 ± 0.01

Table S2. . $^{13}C^{\gamma/\delta}$ relaxation parameters for apo RNaseH

Residue	η_{xy} (s^{-1})	R_1 (s^{-1})
10	3.52 ± 0.39	0.63 ± 0.14
16	2.91 ± 0.22	0.61 ± 0.05
32	3.49 ± 0.42	0.38 ± 0.03
44	2.44 ± 0.11	0.64 ± 0.04
48	3.56 ± 0.49	0.41 ± 0.08
57	4.43 ± 0.17	0.48 ± 0.05
64	2.86 ± 0.09	0.47 ± 0.03
70	4.08 ± 0.64	0.54 ± 0.03
72	0.75 ± 0.14	0.64 ± 0.03
76	1.61 ± 0.36	0.66 ± 0.04
80	0.51 ± 0.29	0.59 ± 0.04
84	1.87 ± 0.14	0.79 ± 0.04
94	1.90 ± 0.27	0.43 ± 0.03
102	5.47 ± 0.83	0.41 ± 0.10
105	3.28 ± 0.55	0.52 ± 0.06
108	5.52 ± 0.57	0.38 ± 0.02
113	0.32 ± 0.11	0.56 ± 0.02
115	0.22 ± 0.15	0.62 ± 0.02
119	1.93 ± 0.09	0.46 ± 0.02
129	3.39 ± 0.42	0.43 ± 0.05
130	6.06 ± 0.63	0.57 ± 0.09
131	1.90 ± 0.24	0.48 ± 0.03
134	3.29 ± 0.19	0.52 ± 0.01
135	1.63 ± 0.13	0.48 ± 0.02
143	0.96 ± 0.11	1.11 ± 0.04
147	0.73 ± 0.06	0.53 ± 0.02
148	4.56 ± 1.54	0.43 ± 0.07
152	1.79 ± 0.07	0.60 ± 0.06
154	0.38 ± 0.04	0.41 ± 0.02

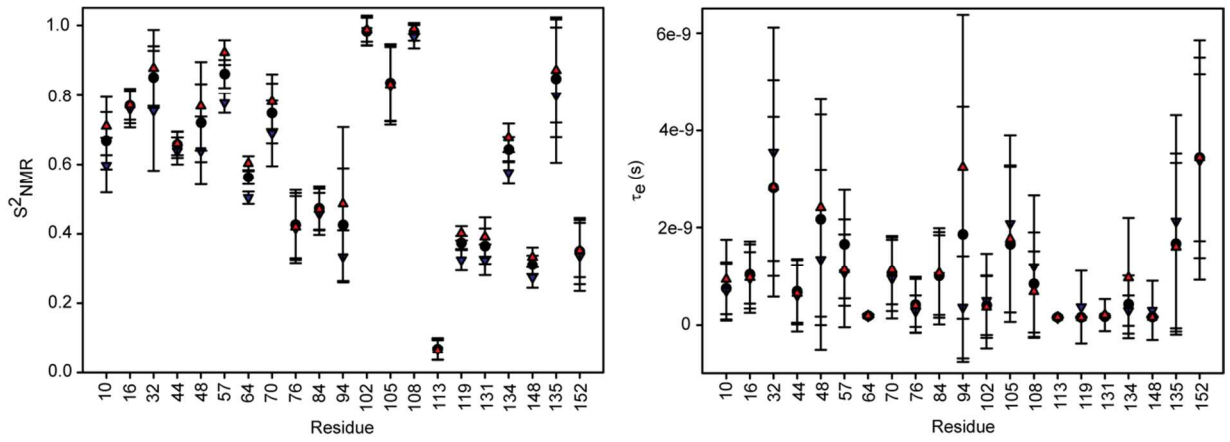

Figure S1. Effect of choice of chemical shift anisotropy (CSA) tensor to (A) order parameters, S^2_{NMR}, and (B) internal correlation time, τ_e. Average CSA of backbone carbonyl 13C measured by solid state NMR[5] were used for Asn and Gln; σxx = 240.87 ppm, σyy = 196.62 ppm σzz = 93.5 ppm. Influence of extreme CSA values to model free analysis were tested; (downward triangles) σxx = 242.87 ppm, σyy = 190.87 ppm σzz = 97.7 ppm (upward triangles), σxx = 243.82 ppm, σyy = 187.87 ppm σzz = 99.3 ppm. This range corresponds to the variation of CSA tensor depending on the change of isotropic chemical shifts observed for backbone carbonyl 13C of GB1. For side chain carboxyl 13C, CSA were assumed to be σxx = 242 ppm, σyy = 191 ppm σzz = 105 ppm,[6] with the σxx component parallel to the C-C bond. The σyy component is strongly affected by the chemical environment, thus two other extreme values were tested with σyy = 168 ppm (upward triangles) and σyy = 202 ppm (downward triangles), while σxx and σzz were remained unchanged.[11]

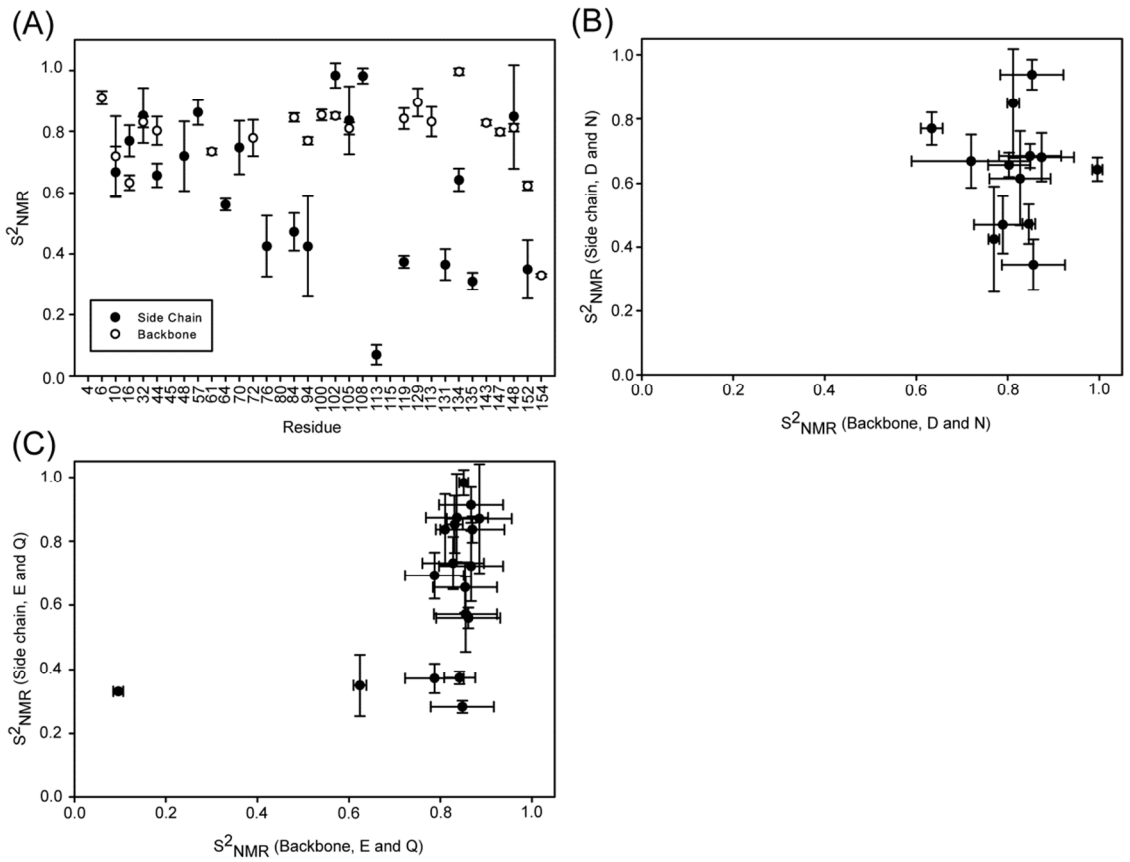

Figure S2. (A) Comparison of order parameters between backbone N-H (open circles) and side chain C-CO (closed circles) bond vectors. (B) Correlation plot of order parameters between backbone and side chain of the D and N residues and (C) E and Q residues. No correlation of the order parameters between backbone and side chain was observed.

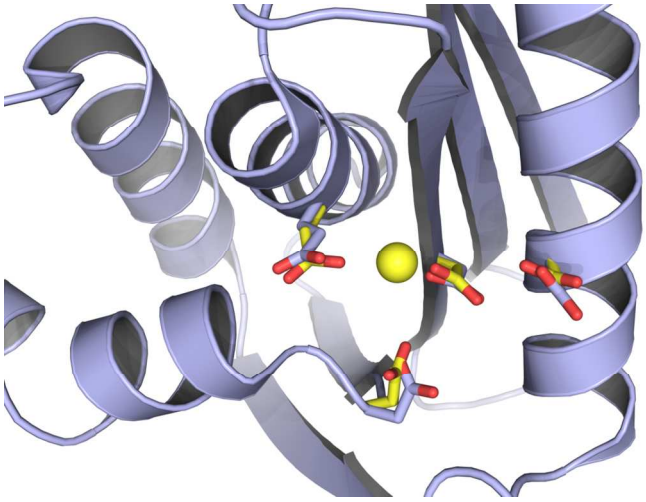

Figure S3. Structural superposition of *E. coli* RNase H in the apo state and in the Mg^{2+} bound state. The apo state (PDB ID 2RN2) is shown in light blue and the Mg^{2+} bound state (PDB ID 1RDD) is shown in yellow. The four active-site carboxylate residues are shown as sticks.

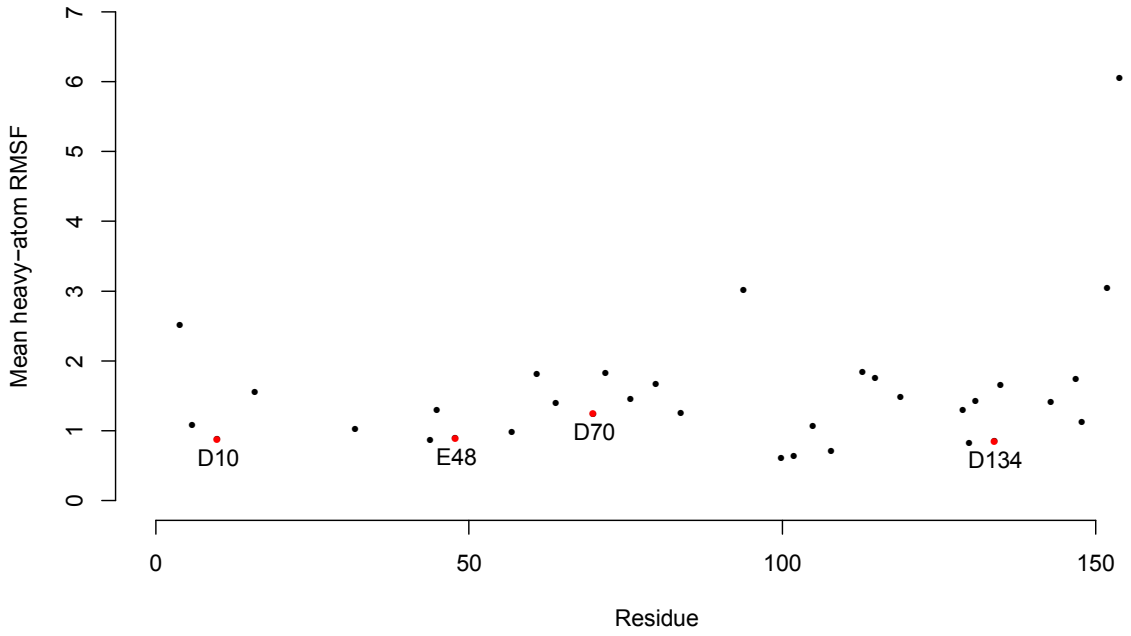

Figure S4. Mean heavy-atom root-mean-square fluctuations (RMSF) for all Asx/Glx side chains. RMSF (Å) was calculated over the 100ns low-pH MD trajectory. Active-site residues are labeled and highlighted in red.

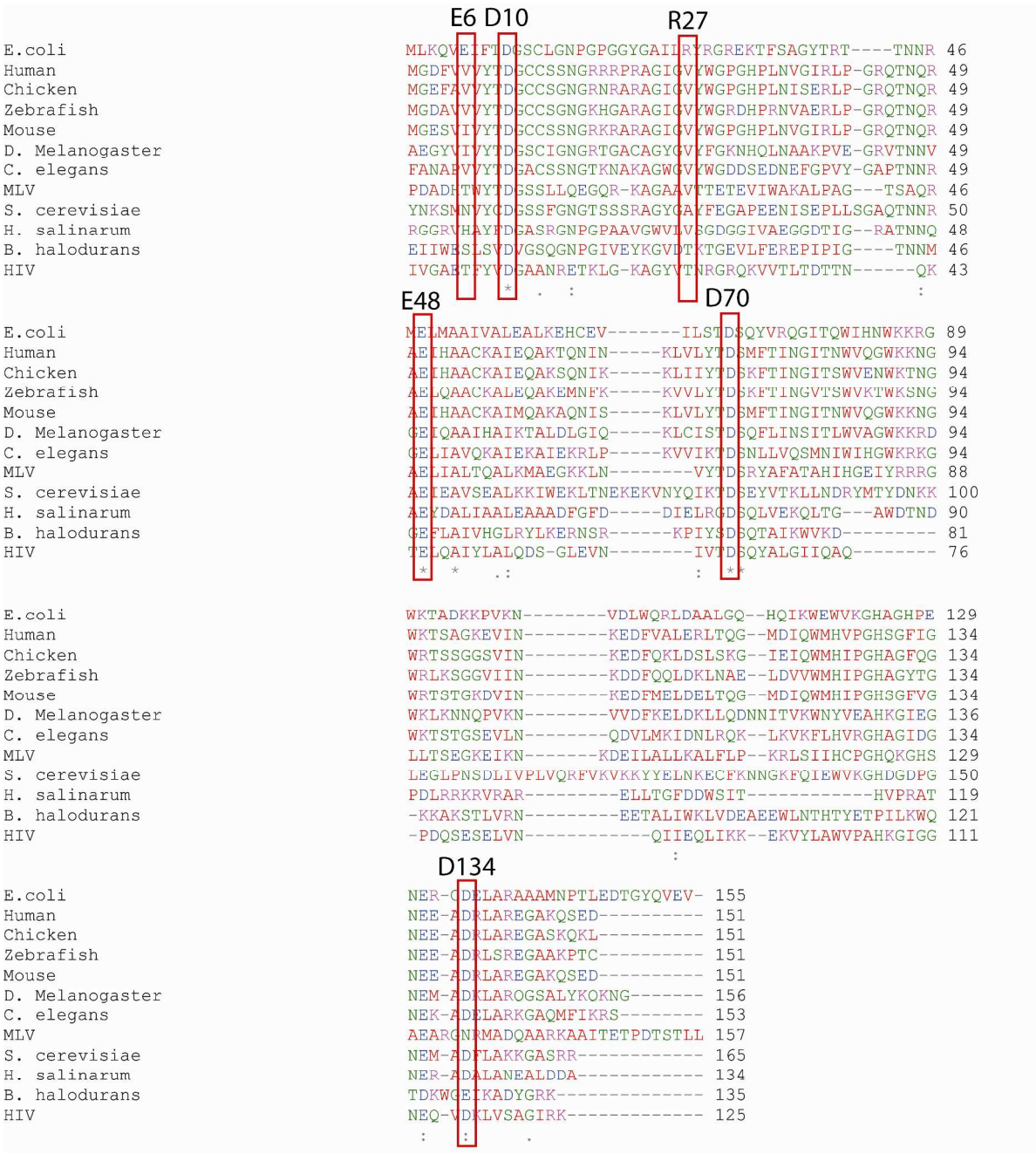

Figure S5. Protein sequence alignment of multiple RNase H domains. Key catalytic residues (D10, E48, D70 and D134) are shown with label. Two partially solvent exposed (relative solvent accessibility ≈ 28 %) salt bridge pair (E6 and R27) is also labeled. This salt bridge is often replaced by hydrophobic residues in other organisms. This might indicate that the salt bridge is formed to reduce the energetic penalty associated with charge burial. Sequence alignment was conducted using ClustalW2.

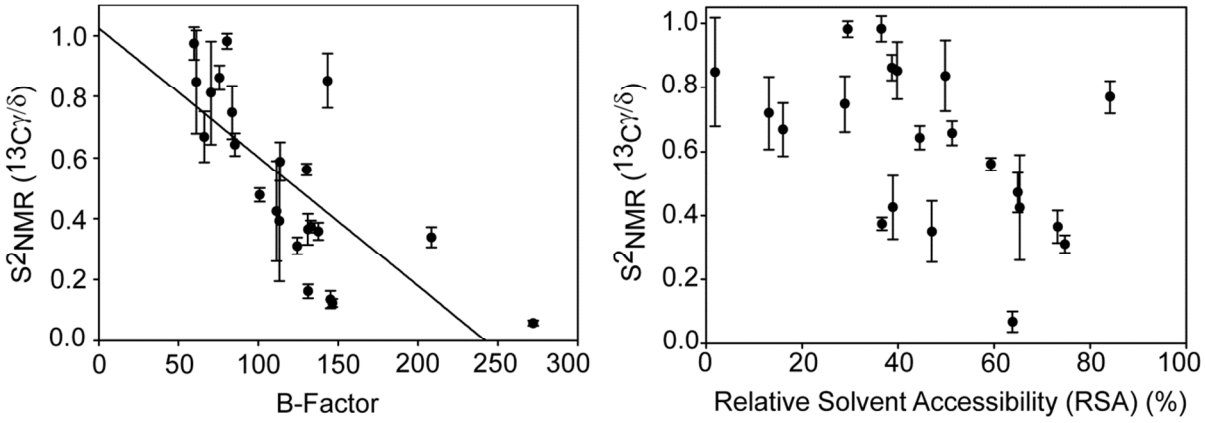

Figure S6. (A) Correlation plot between experimental order parameters, S^2_{NMR}, and crystallographic B-factor (pdb id: 2RN2) for the side chain. Correlation coefficient, R^2 = 0.53 and r.m.s.d. = 0.19. (B) Experimental order parameters vs. RSA of the side chain calculated from the crystal structure of RNase H (pdb id: 2RN2). Correlation coefficient, R^2 = 0.24 and r.m.s.d. = 0.22.

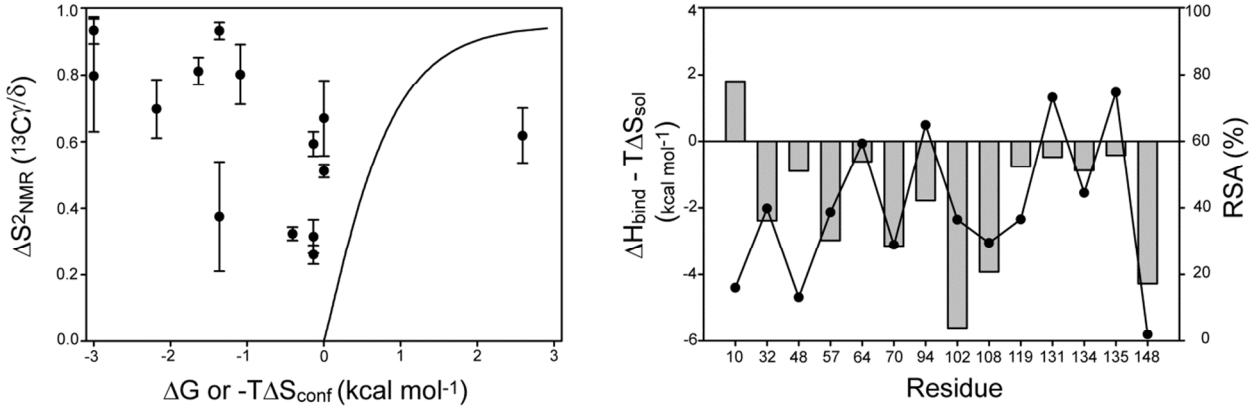

Figure S7. (A) Plot of experimental order parameters vs. free energy change associated with the changes in pK_a = -RTln(ΔpKa) of each acidic residue. Solid line represents the $-T\Delta S_{conf}$.[12] For calculation, order parameter was assumed to change from 0.05 to each S^2_{NMR} (i.e. $\Delta S^2_{NMR} = S^2_{NMR} - 0.05$). (B) Calculated $\Delta H_{bind} - T\Delta S_{sol}$ (bars) and relative solvent accessibility (closed circles) for each acidic residue. Uncertainties for the reported pKa values are not available so the calculated $\Delta H_{bind} - T\Delta S_{sol}$ do not include uncertainties.

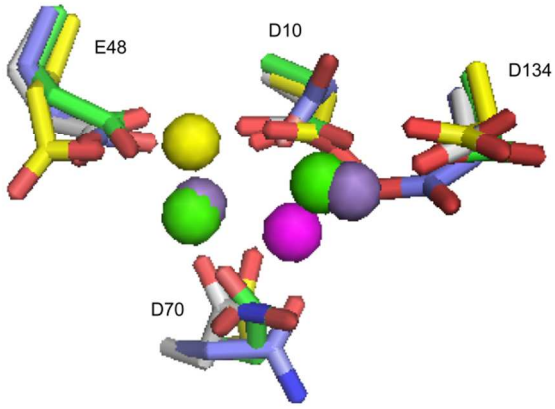

Figure S8. Heterogeneity of metal ion position in the active site of RNase H. Key catalytic residues are shown in stick model and metal ions are shown in sphere; 1RDD (yellow, single Mg^{2+} bound), 2QKK (green, two Ca^{2+} ions bound), 2QKK (magenta, single Ca^{2+} ion bound) and 1G15 (purple, two Mn^{2+} ions bound).

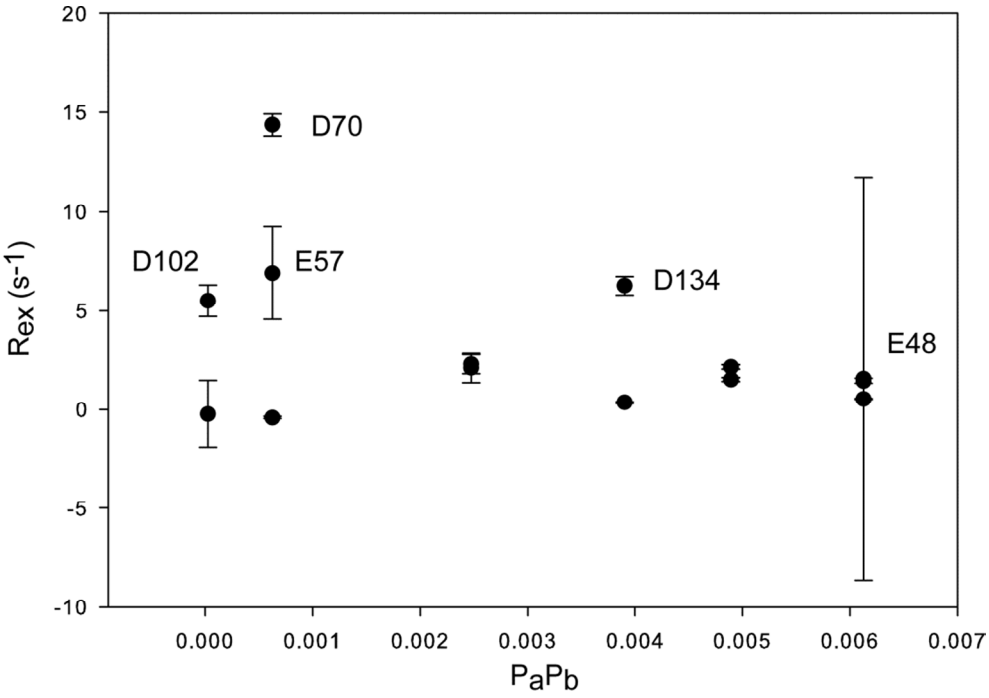

Figure S9. Plot of R_{ex} versus the P_aP_b, the product of major and minor populations. Major and minor populations were calculated by reported pKa values of Mg-bound RNaseH.[13]

References

(1) Paquin, R.; Ferrage, F.; Mulder, F. A.; Akke, M.; Bodenhausen, G. *J. Am. Chem. Soc.* **2008**, *130*, 15805.

(2) Tjiandra, N.; Feller, S. E.; Pastor, R. W.; Bax, A. *J. Am. Chem. Soc.* **1995**, *117*, 12562.

(3) Schurr, J. M.; Babcock, H. P.; Fujimoto, B. S. *J. Magn. Reson. B* **1994**, *105*, 211.

(4) Kroenke, C. D.; Loria, J. P.; Lee, L. K.; Rance, M.; Palmer, A. G. 3rd. *J. Am. Chem. Soc.* **1998**, *120*, 7905.

(5) Wylie, B. J.; Sperling, L. J.; Frericks, H. L.; Shah, G. J.; Franks, W. T.; Rienstra, C. M. *J. Am. Chem. Soc.* **2007**, *129*, 5318.

(6) Gardiennet-Doucet, C.; Henry, B.; Tekely, P. *Prog. Nucl. Magn. Reson. Spectrosc.* **2006**, *49*, 129.

(7) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roiterg, A.; Simmering, C. *Proteins: Struct. Funct. Bioinf.* **2006**, *65*, 712.

(8) Bowers, K. J.; Chow, E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D.; Salmon, J. K.; Shan, Y.; Shaw, D. E. In *Proceedings of the 2006 ACM/IEEE conference on Supercomputing*; ACM: Tampa, Florida, 2006, p 84.

(9) Chandrasekar, I.; Clore, G. M.; Szabo, A.; Gronenborn, A. M.; Brooks, B. R. *J. Mol. Biol.* **1992**, *226*, 239.

(10) Stafford, K. A.; Robustelli, P.; Palmer, A. G. 3rd. *PLoS Comput. Biol.* **2013**, *9*, e1003218.

(11) Gu, Z.; Zambrano, R.; McDermott, A. *J. Am. Chem. Soc.* **1994**, *116*, 6368.

(12) Yang, D.; Kay, L. E. *J. Mol. Biol.* **1996**, *263*, 369.

(13) Oda, Y.; Yamazaki, T.; Nagayama, K.; Kanaya, S.; Kuroda, Y.; Nakamura, H. *Biochemistry* **1994**, *33*, 5275.